Wafer Bumping Process and Inter-Chip Connections for Ultra-High Data Transfer Rates in Multi-Chip Modules With Superconductor Integrated Circuits

Sergey K. Tolpygo, Diana Tolpygo, Richard T. Hunt, Supradeep Narayana, Yuri A. Polyakov, and Vasili K. Semenov

Abstract—**Josephson junction logic cells and superconductor microstrip lines are able to process and transfer digital data with rates up to several hundred GHz as has been demonstrated in single-chip experiments. However, the existing chip-level bumping technique in InSn solder and resulting inter-chip connections do not allow expanding these rates to multi-chip circuits. We developed a wafer-level bumping technology using lithographically-defined bumps deposited either by e-beam evaporation or electroplating, and proposed and implemented a novel design of high-frequency chip interconnects. Chip-to-chip single-flux-quantum pulse transmission rates reaching 110 GHz have been achieved. The observed rates were limited not by the interconnects but by the speed of on-chip test circuitry fabricated in the framework of 4.5 kA/cm^2 HYPRES process for superconductor integrated circuits. Experimental results on adhesive-bonded and reflow-bonded multi-chip modules (MCMs) with Au and InSn bumps are presented, and effective parameters of the new interconnect design and MCM technology are discussed.**

Index Terms—**Flip-chip devices, integrated circuit interconnections, multichip modules, superconducting integrated circuits.**

I. INTRODUCTION

IT is well known that Josephson-junction-based logic cells and superconducting microstrip lines (MSLs) are able to process and transfer digital data with rates up to several hundred GHz. Relatively complex superconductor integrated circuits (SICs) with clock frequencies of ∼30 GHz have recently been demonstrated. The complexity and functionality of SICs are limited by the logic cell density or, crudely, by the number of Josephson junctions (JJs) which can be placed on a single chip. For the existing IC fabrication technologies (see, e.g., HYPRES 4.5 kA/cm^2 Josephson critical current density process [1]), the maximum circuit density is ∼ $2 \cdot 10^4$ JJs per cm^2 and mainly restricted by linewidth of the employed photolithography tools, whereas the active chip size ∼ 1 cm^2 is limited by the fabrication yield (density of defects). The clock frequency of SICs can be increased further by simply increasing the Josephson current density j_c and an appropriate scaling of junctions' areas [2]. Increasing of chip functionality requires however an overall scaling down of logic cell dimensions and increasing the number of active superconducting layers [3], [4]. Similarly to semiconductor technologies, a natural way of building more complex superconducting circuits and systems is to use multichip module (MCM) technologies [5]–[7]. To be useful, this requires transferring data between chips in the MCM at the same or similar rates as on-chip, i.e. up to several hundred Gbit/s and beyond, and can only be achieved in a flip-chip configuration.

There have been a number of publications on flip-chip technology for superconductor MCMs, interconnect design, and experiments on data transfer rates in flip-chip MCMs [8]–[13]. They all employed the most primitive version of the original control collapse chip connection (C4) process, developed at IBM 50 year ago [14], in which each chip and/or the MCM carrier are manually dipped into a molten solder to form solder bumps on bump-bonding pads having an appropriate underbump metallization. The chips are then attached to the carrier by a flip-chip bonder using either a reflow process or an adhesive [15]. The original Sn-Pb alloy was replaced in this process by a low melting point InSn eutectic because superconductor circuits (Nb/AlO$_x$/Nb JJs) usually irreversibly change their properties when heated above ∼ 180°C. In our experience, this manual immersion is very laborious, low yield, and low throughput process requiring multiple reworks in order to achieve full wetting of bump-bonding pads and acceptable uniformity of bump heights. Another big disadvantage of this immersion process is that, because of the nature of wetting, the height of solder bumps depends on their diameter so they cannot be changed independently [16], [17]. The higher is the desired data transfer rate the shorter should be the bump (see below). Therefore, increasing the data rates and the number of I/Os requires wetting smaller and smaller contact pads, which makes the immersion process yet more difficult and unreliable. The existing designs of superconductor chip interconnects and their electric model are not much different from those proposed
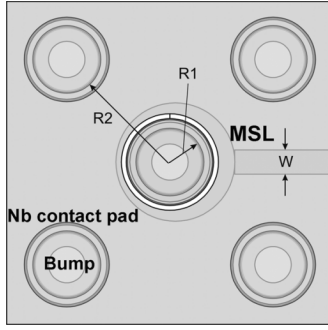
Fig. 1. Structure of a wide-band chip-to-chip connection. Signal bump (in the center) is surrounded by 4 ground bumps. Bumps are deposited into the centers of Nb contact pads. Superconducting microstrip line (**MSL**) is connected to the signal bump. It has a specially shaped end overlapping a cut in the ground plane in order to provide impedance matching and increase the transmission bandwidth.

30 years ago [18], and present a peripheral array of signal and ground bumps connecting superconducting microstrip lines on flip chips and on the MCM carrier.

In this work, instead of manual chip bumping, we developed a whole wafer bumping technology using lithographically-defined bumps deposited either by e-beam evaporation or electroplating. This provides a 100-fold increase in the process throughput and over a 50-fold increase in the bump uniformity as well as enables getting practically any desired height, diameter, and number of the bumps. We also proposed a novel design of interconnects which allowed us to achieve chip-to-chip data transfer rates in excess of 110 GHz in the frameworks of HYPRES 4.5 kA/cm^2 fabrication technology.

II. CHIP INTERCONNECTS PROPERTIES AND DESIGN

The usual arrangement of signal and ground contacts for chip-to-chip communication is a peripheral array shown in Fig. 1. Each signal contact is surrounded by four ground contacts which can also be shared with the next signal contact. The signal (data) is coming along a microstrip line (width w) formed in one of the superconducting layers and marked **MSL** in Fig. 1. The bumps are formed in the centers of the signal and ground contact pads. Their diameter is somewhat smaller that the diameter of the contact pads. The MSL impedance is matched to a load or a driver at the other end of the MSL [19].

Conventional solder bumps have relatively large parasitic inductance L_p that is acceptable for relatively slow semiconductor devices having a high, typically $R = 50\ \Omega$, impedance. However, high parasitic inductance is not acceptable for a low-impedance, typically $\sim 2\ \Omega$, superconductor devices operating at about 100 GHz data rate. This is because the bump inductance limits the cutoff frequency for signal transmission through the bump, ω_{cf}, which can be crudely estimated as

$$\omega_{cf} \sim R/L_p. \tag{1}$$

The parasitic inductance can be reduced by reducing the contact pad size $R1$ (Fig. 1). This approach however has evident technological limitations if the bumps are formed by manual solder wetting. Besides, this reduction is rather small because

it is proportional only to $\ln(R1)$. Another known approach is to compensate the series parasitic inductance by a proper parallel capacitor [18]

$$C_c = L_p/R^2. \tag{2}$$

However, the cutoff frequency for this solution is limited by the resonant frequency, giving the same answer as (1)

$$\omega_r = 1/\sqrt{L_p \cdot C_c} = R/L_p. \tag{3}$$

Besides, this technique works only if the parasitic inductance is known with a relatively high accuracy.

Our new interconnect design actively utilizes advantages of several metallization levels common for superconductor integrated circuit technology. The ground plane bumps occupying corners in Fig. 1 provide galvanic connections between the ground planes of all chips comprising the MCM. Their contact pads contain all superconducting layers (from M0 to M3 in the HYPRES process) mutually connected via contact holes in isolating layers. The sequence of metal layers in the signal contact pad is identical to those of the ground contacts. However, the patterns of the layers and therefore the bump properties are original. First, the signal contact pad is isolated from the ground plane by a narrow disk-shaped moat shown in white. The microstrip line is galvanically connected to the bump. Second, the bump end of the MSL is shaped such that it overlaps the ground plane moat and behaves as a line with gradually reducing width. One of the possible implementations of this transition is a "misaligned" circular end shape formed in the same M2 layer as the microstrip, as shown in Fig. 1.

The complementary signal bump located on the MCM carrier has exactly the same design but is just mirrored with respect to the y-axis. As a result, the whole structure after flip-chip bonding behaves almost as a uniform microstrip line which has a transition in the z-direction from the MCM carrier (base chip) onto the flipped chip. Of course the transition creates a parasitic inductance in the vertical z-direction, which for this geometry is small and can be estimated as

$$L_p = L_W \cdot N, \tag{4}$$

where the sheet inductance $L_W = \mu_0 \cdot t_m$ and the number of squares

$$N = \ln(R2/R1)/2\pi. \tag{5}$$

Magnetic gap $t_m = t + 2\lambda$, where λ is magnetic field penetration depth, is the only parameter depending on the spacing, t, between the substrate and the flip chip, i.e. on the bump height. It crudely equals to the distance between the ground planes on the MCM carrier and on the flip-chip. At $t_m = 3\ \mu m$, $L_W \approx 3.8$ pH.

Another convenient feature of the suggested design is that N depends only on $R2/R1$ ratio and does not depend on the contact pad and bump diameters. At $R2/R1 = 5$ the number of squares equals to 0.25. As a result, the parasitic inductance is about 1 pH. Even for a very low (e.g., $2\ \Omega$) impedance of the MSL, the characteristic resonant frequency ω_r according to (3)

TABLE I
UNIFORMITY OF LITHOGRAPHYCALLY-DEFINED BUMPS

Location	Average bump height, h_{av} (nm)	Uniformity $(h_{max}\text{-}h_{min})/2h_{av}$
Across 150-mm wafer	1590	± 2.03%
Across 5x5 mm² chip, central part of the wafer	1623	± 0.16%
Across 5x5 mm² chip, worst location near the wafer edge	1565	± 0.26%
Across 10x10 mm² chip, worst location near the wafer edge	1568	± 0.59%

is about $2 \cdot 10^{12}$ corresponding to the cutoff frequency of about 300 GHz. In practice, the cutoff frequency can be even higher. This is because the parasitic inductance L_p is not a lump element but is distributed along the microstrip line with varying width. The distributed inductance can be compensated by a distributed capacitance. In the design, this compensation is achieved by widening the MSL's end and increasing the overlap with the ground plane. The interconnect geometry shown in Fig. 1 includes this capacitive compensation. We adjusted the M2/MO overlap area at a fixed $R1$ to achieve the bandwidth larger than 150 GHz. For the highest possible bandwidth, both the direct measurements and more detailed analysis should be carried out to ensure that the geometry is completely optimized.

Yet another advantage of the proposed design is that the signal contact pad, although containing the ground plane layer, is isolated from the circuit ground in the $x - y$ plane by a moat rather than by the interlayer dielectric in the z-direction as is common to many conventional designs. This makes it possible to apply significant pressure to the bump during flip-chip bonding without punching through the insulation and shorting the signal bump to the ground. Also the distributed capacitance can be varied independently of the size of contact pads and bumps.

III. FABRICATION PROCESS

A number of test circuits for measuring data transfer rates between superconductor chips were designed using the contact pad diameter of 75 μm and the bump diameter of 30 μm. The circuits were fabricated by 11-layer HYPRES process using Nb/Al/AlO$_x$/Nb junctions with 4.5 kA/cm² current density. A large number of wafers have been fabricated for testing different bump metallurgies, deposition methods, and flip-chip bonding schemes. Bumps for flip-chip bonding were defined all over the 150-mm process wafers using a lift-off process. The resist overhang profile for the lift-off was achieved by using AZ-5214E or AZ-5218E photoresist in the image reversal mode and a clear field photomask. For the lift-off process, the bumps were deposited by e-beam evaporation on top of Nb contact pads with Ti/Pd adhesion layer. Wafers with three types of bump material were fabricated: Au, Cu, and In. The total height of the bumps was chosen to be in the range from 1.6 μm to 2.0 μm. To prevent Cu and In bumps from oxidation during wafer storage before flip-chip bonding, a capping layer of 50-nm-thick Au was deposited in the same process. Thicker bumps were deposited by electroplating onto e-beam evaporated Ti/Pd/Au underbump metallurgy. For a comparison, some
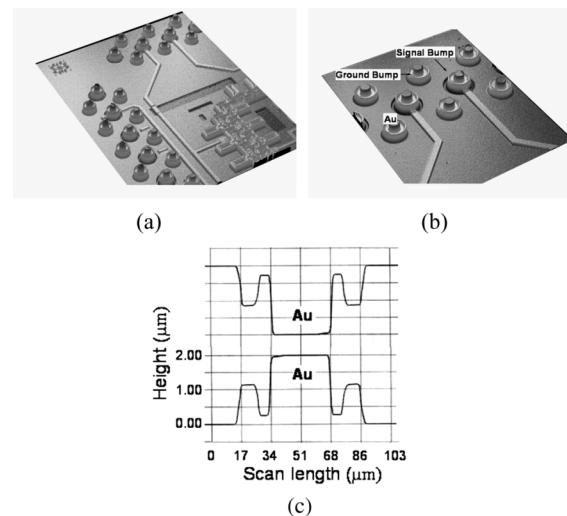

(a) (b)

(c)

Fig. 2. (a) 3D image of a part of the fabricated chip, showing the array of ground bumps (free standing, not connected), which surround the signal bumps connected to bias leads and microstrip lines; (b) blow-up of the bump area; (c) cross-section scan of the ground bumps and their position in a flip-chip configuration just before making contact. Each soft gold bump in the center is surrounded by a hard ring of a smaller height that is designed to serve as a compression limiter (stopper), see also Fig. 2(b).

of the test chips were bumped using the manual immersion into molten InSn eutectic alloy at 150°C.

The fabricated wafers have been characterized using both a stylus and an optical profilometers in order to evaluate bumps uniformity and shape. For instance, Table I summarizes results for Au bumps. We can see that the process produces very uniform bumps all across 150-mm wafers. Even in the worst part of the wafer near the very edge, the maximum total difference in the bump heights on a 1×1 cm² chip is only 18 nm. It means that very little compression will be needed to achieve a good electrical contact between all the bumps in a flip-chip MCM structure. For a comparison, the best uniformity we were able to achieve with the manual immersion process was ±15% for 7-μm InSn bumps on 5×5 mm² chips and yet worse for smaller heights required for ultrahigh data transfer rates. Therefore, a more than 50-fold improvement in the bump height uniformity was achieved by implementing e-beam evaporated or electroplated bumps. Very similar results on bumps uniformity were obtained with Ti/Pd/Cu/Au and T/Pd/In/Au bumps.

The typical surface scan across the ground bump is shown in Fig. 2. The total bump height (bump + underlayers) measured from the field oxide surface over the ground plane is 1.99 μm as shown in Fig. 2. This results in $\approx 4 - \mu$m separation between the flip-chip and the MCM substrate (the distance from the field oxide on the chip to the field oxide on the substrate). The tallest parts of the circuits in the HYPRES fabrication process can stick out of the field oxide by as much as 1.09 μm. It means that the typical separation between the top Nb layer (M3 layer) on the flipped chip and on the MCM substrate is 1.8 μm when opposing bumps are brought into contact. This separation provides enough room for compressing the bumps in order of achieving good electric contacts between the opposing bumps. In order to avoid over-compression and complete flattening of the soft indium or gold bumps, each bump was designed such that it is
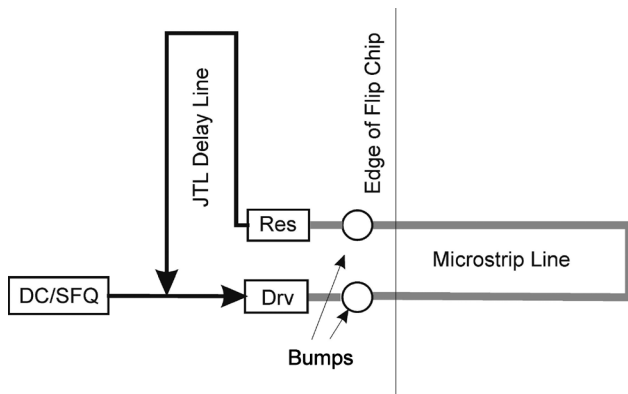
Fig. 3. Block diagram of the ring oscillator used for testing SFQ pulse transmission between the flip chip and the active MCM carrier. SFQ driver and receiver are marked **Drv** and **Res**, respectively.
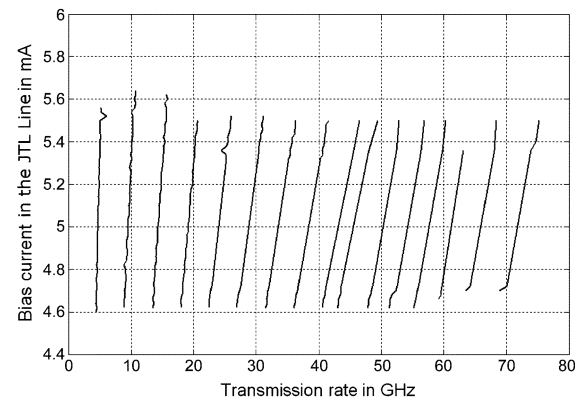

Fig. 4. The frequency of transmission of SFQ pulses between the flip-chip and the MCM carrier in the ring oscillator at various numbers of flux quanta circulating in the ring for the typical MCM. The maximum of 16 SFQ pulses can be inserted into the ring oscillator. The speed of circulation of the pulses was controlled by varying the bias on the JTL.

surrounded by a ring made of much harder layers of Nb and SiO_2 (see Fig. 2). This ring serves as a compression limiter preventing the bump material to be squeezed out and short to the nearest bump.

All MCMs were assembled using an FC-150 flip chip bonder from *Suss Microtec*. Several bonding schemes were tested such as compression and reflow for low melting point In and InSn bumps, compression plus adhesive bonding for Cu, Au, and InSn bumps, and reflow plus adhesive underfill for increasing mechanical strength, all giving essentially the same results in MCM testing. A cryogenically compatible adhesive was chosen such that it provides additional compression of bumps upon cooling due to the difference in coefficients of thermal expansion.

IV. TEST RESULTS

Test circuits were placed on both the flip chips and the MCM carriers. Their block diagram is shown in Fig. 3. The circuits presented ring oscillators for testing transmission of single flux quantum (SFQ) pulses from the Josephson transmission line (JTL) on the chip through the bumps to the microstrip line on the carrier and back to the chip. Similarly, SFQ pulses generated on the carrier can be transmitted to the flip chip and back to the carrier. The length of the microstrip lines in different circuits was varied from 300 μm to 13 mm in order to see if there are any resonances or other limitations on the data rates associated with the length of the microstrips. The speed of the pulse circulation was controlled by varying the bias on the JTL as well as by the number of flux quanta inserted into the ring. For a comparison, identical circuits were placed entirely on the chip and on the carrier, so the SFQ pulses do not need to propagate through the bumps. Testing was done using an automated set-up *Octopux*.

Fig. 4 shows the frequency of transmission of the SFQ pulses through the bumps in the ring oscillator at different biases of the JTL controlling the SFQ pulse propagation speed. Each curve from left to right in Fig. 4 corresponds to a progressively increasing number of SFQ pulses moving in the ring from one to sixteen.

Fig. 5 shows the margins of operation of the receiver (**Res** in Fig. 3) in the ring oscillator as a function of SFQ pulse transmission rate in the ring. The receiver is the most sensitive part

Wafer KL1061, bumped by evaporated 1.8-μm Au						
MSL length (μm)	3580	3680	3760	3845	4315	7050
f_{max} (GHz)	104	108	108	109	88	105
Wafer KL1084, bumped by immersion in InSn melt						
MSL length (μm)	1900	2210	2335	4475	7650	12895
f (GHz)	78	80	78	76	81	76
f (GHz)	78	80	78	72	81	72

Fig. 5. The upper and lower margins of the receiver bias current at different rates of SFQ pulse transmission between the flip-chip and the MCM carrier with Au bumps. The MSL length is 2210 μm. The wafer was bumped using 1.8 μm Au bumps deposited by e-beam evaporation. The MCM was bonded by a cryogenically compatible adhesive.

of the whole circuit, so the margins on its bias current are the most representative of the circuit operation. The maximum SFQ pulse transmission rate, f_{max} observed is about 110 GHz.

Fig. 6 shows the results for the same circuit but from a wafer bumped manually by the traditional immersion process. The $f_{max} \sim 78$ GHz in this MCM is somewhat lower than in the previous one with the Au bumps. However, by testing the ring oscillator located entirely on the flip-chip of this MCM, we found basically the same maximum transmission rate of SFQ pulses. It means that the observed f_{max} is limited not by the transmission through the bumps but by the operation speed of the circuit itself which varies somewhat from wafer to wafer. The reproducibility of the performance of different circuits on the same MCM is shown in Table II for the two typical MCMs from two different wafers. The circuits differ by the length of the microstrip lines.

In some rare cases with InSn bumps we observed a reduction of margins of the circuit operation in some intermediate range of frequencies. We attribute this to a resonance in the MSL that can occur due to some imperfection of particular bumps or their bonding. The typical example is shown in Fig. 7 displaying a broad resonance at ~44 GHz.

V. CONCLUSION

We have developed a new chip interconnect design and the wafer-level bumping process for superconductor integrated circuits, using evaporated or electroplated bumps of various com-
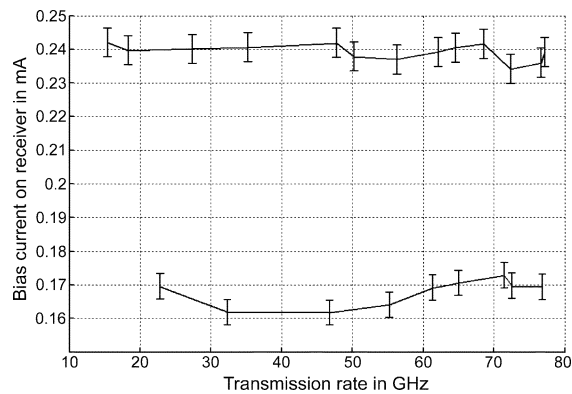
Fig. 6. The upper and lower margins of the receiver bias current vs the SFQ pulse transmission rate between the flip-chip and the MCM carrier for InSn bumps. Both the chip and the carrier were bumped by immersion in molten InSn alloy at 150°C and bonded by reflow and a cryogenically compatible adhesive. The MSL length is 2210 μm.

TABLE II
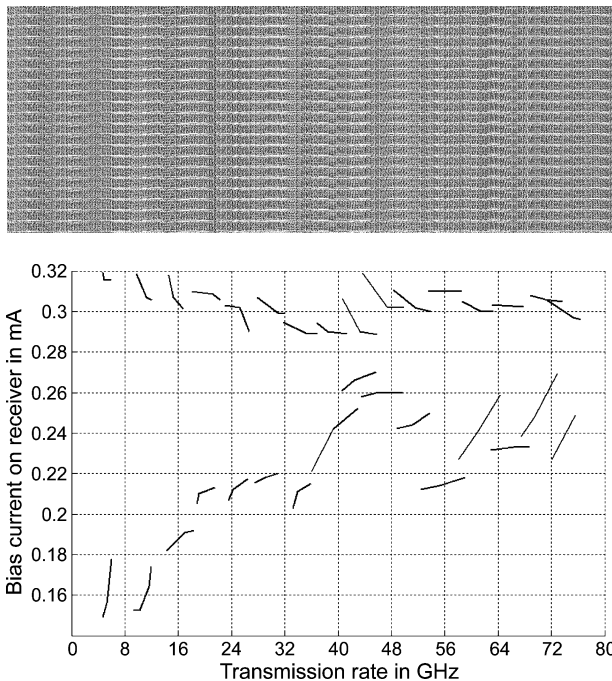THE MAXIMUM DATA TRANSER RATES

Fig. 7. The upper and lower margins of the receiver bias current vs the SFQ pulse transmission rate between the flip-chip and the MCM carrier for InSn bumps. Each separate branch from left to right corresponds to the different number of SFQ pulses circulating in the ring oscillator. A resonance at ∼44 GHz can be seen. The MSL length is 2210 μm.

positions—Ti/Pd/Au, Ti/Pd/Cu/Au, Ti/Pd//In/Au—in order to replace the manual chip-level immersion bumping in molten InSn. A large number of MCMs comprised of a single flip-chip and an active MCM carrier have been assembled using adhesive and reflow bonding processes. The maximum chip-to-chip SFQ pulse transmission rates of 110 GHz was observed for the circuits fabricated by HYPRES 4.5 $\mathrm{kA/cm^2}$ process and using evaporated Ti/Pd/Au bumps of 1.8 μm height. By comparing the on-chip (no bumps) and chip-to-chip (through the bumps) SFQ pulse transmission, we found that the maximum observed

frequency was not limited by the employed interconnect design (or the bump metallurgy) but reflected the maximum speed of the RSFQ test circuitry. Therefore, the developed design of interconnects and the whole wafer bumping technology should be well suited for the future higher-j_c fabrication processes providing yet higher operating frequencies of superconductor integrated circuits.

REFERENCES

[1] HYPRES Nb Process Design Rules (30-1000-4500 $\mathrm{A/cm^2}$), Process #03-10-45 HYPRES, Inc.. Elmsford, NY, 10523 [Online]. Available: http://www.hypres.com/pages/download/designrules/DesignRules.pdf

[2] S. K. Tolpygo, D. Yohannes, R. T. Hunt, J. A. Vivalda, D. Donnelly, D. Amparo, and A. F. Kirichenko, "20 $\mathrm{kA/cm^2}$ process development for superconductor integrated circuits with 80 GHz clock frequency," *IEEE Trans. Appl. Supercond.*, vol. 17, pp. 946–951, June 2007.

[3] L. A. Abelson and G. L. Kerber, "Superconductor integrated circuit fabrication technology," *IEEE Proceedings*, vol. 92, pp. 1517–1533, Oct. 2004.

[4] M. Hidaka, S. Nagasawa, K. Hinode, and T. Satoh, "Improvements in fabrication process for Nb-based single flux quantum circuits in Japan," *IEICE Trans. Electron.*, vol. E91-C, pp. 318–324, Mar. 2008.

[5] *Flip Chip Technologies*, J. H. Lau, Ed. Boston: McGraw-Hill, 1996.

[6] *Multichip Module Technology Handbook*, P. E. Garrou and I. Turlik, Eds. New York: McGraw-Hill, 1998.

[7] J. H. Lau, *Low Cost Flip Chip Technologies*. New York: McGraw-Hill, 2000.

[8] K. E. Yokoyama, G. Akerling, A. D. Smith, and M. Wire, "Robust superconducting die attach process," *IEEE Trans. Appl. Supercond.*, vol. 7, pp. 2631–2634, June 1997.

[9] Q. P. Herr, A. D. Smith, and M. S. Wire, "High speed data link between digital superconductor chips," *Appl. Phys. Lett.*, vol. 80, pp. 32310–32312, Apr. 2002.

[10] Q. Herr, M. Wire, and A. Smith, "Ballistic SFQ signal propagation on-chip and chip-to-chip," *IEEE Trans. Appl. Supercond.*, vol. 13, pp. 463–466, Jun. 2003.

[11] Y. Hashimoto, S. Yorozu, and T. Miyazaki, "Transmission of single-flux-quantum pulse between superconductor chips," *Appl. Phys. Lett.*, vol. 86, pp. 072502-1–072502-3, Feb. 2005.

[12] Y. Hashimoto, S. Yorozy, T. Satoh, and T. Miyazaki, "Demonstration of chip-to-chip transmission of single-flux-quantum pulses at throughputs beyond 100 Gbps," *Appl. Phys. Lett.*, vol. 87, pp. 022502–022504, July 2005.

[13] Y. Hashimoto, S. Yorozu, and Y. Kameda, "Development of cryopackaging and I/O technologies for high-speed superconductive digital systems," *IEICE Trans. Electron.*, vol. E91-C, pp. 325–332, Mar. 2008.

[14] L. F. Miller, "Controlled collapse reflow chip joining," *IBM J. Res. Develop.*, vol. 13, pp. 239–250, May 1969.

[15] S. B. Kaplan, V. Dotsenko, and D. Tolpygo, "High-speed experimental results for an adhesive-bonded superconducting multi-chip module," *IEEE Trans. Appl. Supercond.*, vol. 17, pp. 971–974, June 2007.

[16] L. Li, S. Wiegele, P. Thompson, and R. Lee, "Stencil printing process development for low cost flip chip interconnect," in *Proceedings of IEEE Electronic Components and Technology Conf.*, May 1998, pp. 421–426.

[17] M. Cho, S. Kang, Y. Kwon, D. Jang, and N. Kim, "Flip-chip bonding on PCB with electroless Ni-Au and stencil printing solder bump," in *Proceedings of SMTA Intern. Conf.*, Sep. 1999, pp. 159–164.

[18] H. C. Jones and D. J. Herrell, "The characteristics of chip-to-chip signal propagation in a package suitable for superconducting circuits," *IBM J. Res. Develop.*, vol. 24, pp. 172–177, Mar. 1980.

[19] S. V. Polonsky, V. K. Semenov, and D. F. Schneider, "Transmission of single-flux-quantum pulses along superconducting microstrip lines," *IEEE Trans. Appl. Supercond.*, vol. 3, pp. 2598–2600, Mar. 1993.